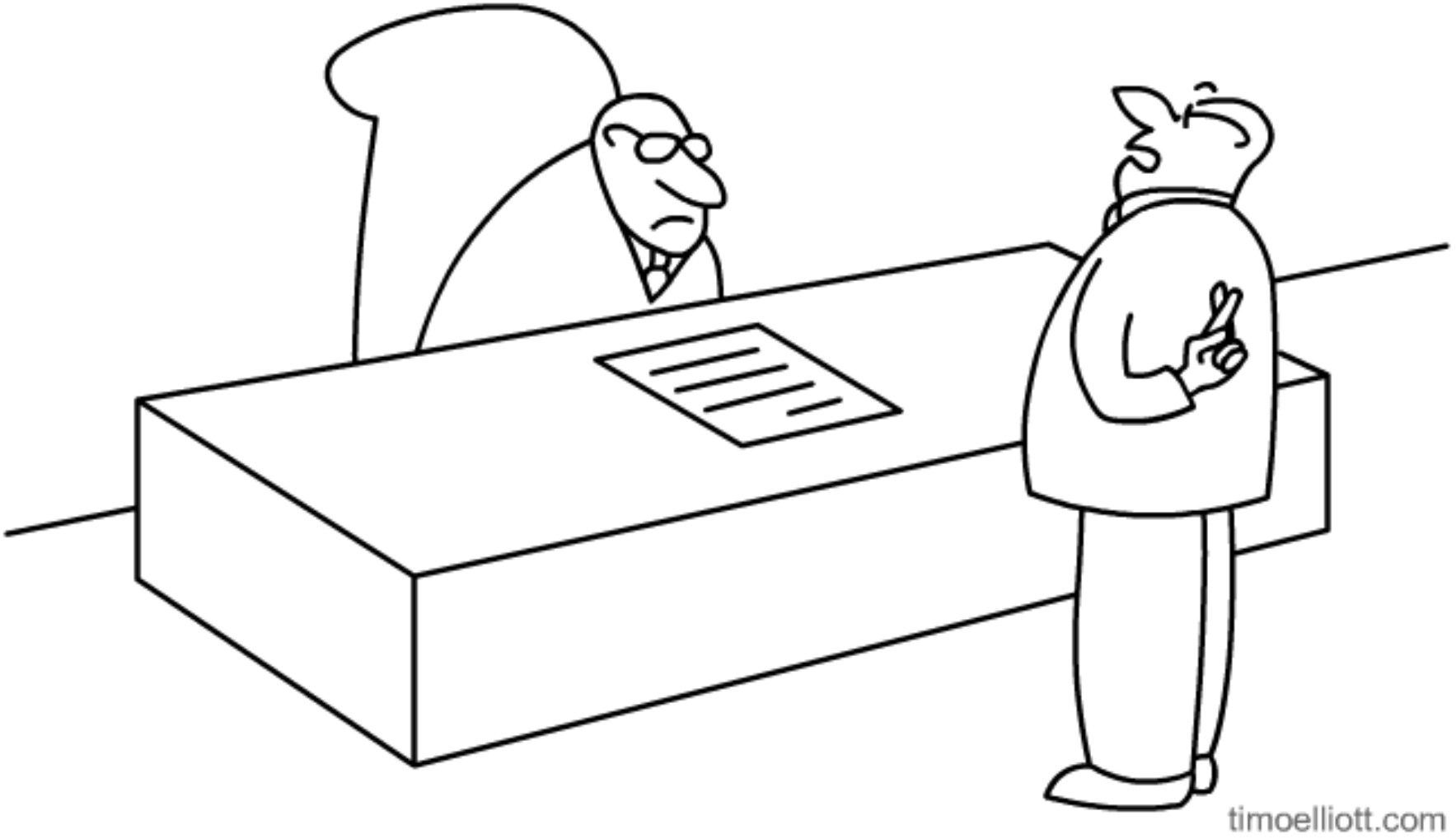
Roadmap to Data Integrity: Practical Data Validation, Verification, and Security Controls

Emory Internal Audit
Scott Stevenson, Chief Audit Officer
Deepa Pawate, Associate Chief Audit Officer
Kristy Habib, Manager, University Internal Audit

What is Data Integrity?



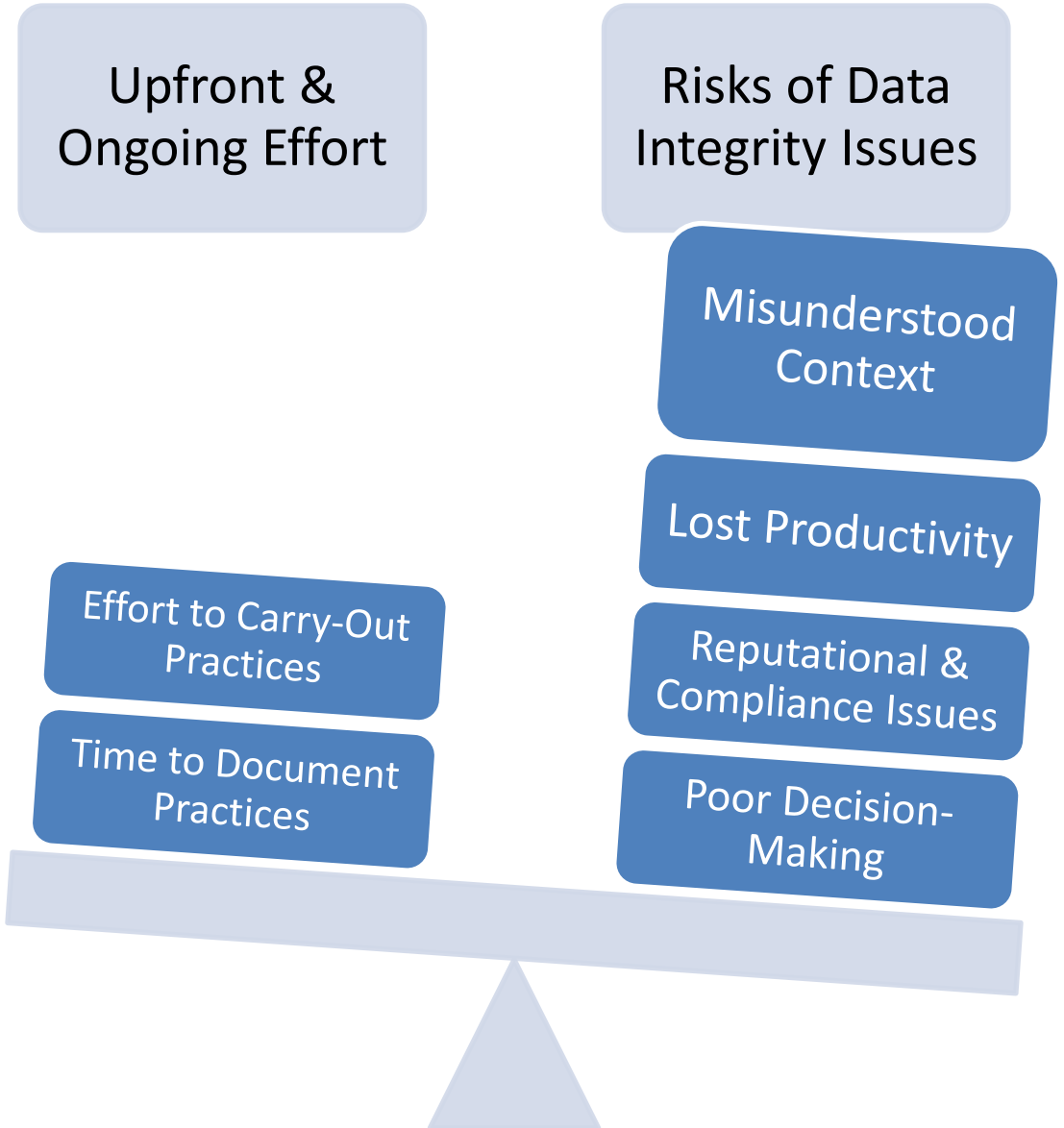
"Yes sir, you can absolutely trust those numbers"

Data Integrity – Multiple Dimensions

Emory Policy – Institutional Data Management says that "data integrity" is composed of:

- **Accuracy:** *Data free from errors*
- **Completeness:** *All values are present.*
- **Consistency:** *Data satisfy a set of definitions or constraints that applied and maintained in the same manner across reports*
- **Reliability:** *Independent custodians or users obtain consistent results when applying the same definitions or constraints.*
- **Timeliness:** *Data are available when required.*

Why Does Data Integrity Matter?



Examples of Data Integrity & Quality Issues

Promoting Data Integrity = Validation + Verification + Security

(1) Data validation – *"The computer checks if the data is **correct and makes sense** (e.g., correct format, length, etc.)."*

(2) Data verification and approval – *"People step-in to confirm that the data is **reliable and accurate** (e.g., reported data matches source data, etc.)."*

(3) Data security – *"People limit access to data sets, systems, and spreadsheets to only those **with a need to know** (e.g. formal user access request process, access terminations, audit logs)."*

Data Integrity – Key Controls for Your Unit/Department

(1) Document your unit's data management procedures and controls:

- Data validation,
- Data verification, and
- Data security

** Include source systems for data, where data is downloaded into (e.g., spreadsheet), who has access, how data is manipulated/edited to meet context of questions being asked, who reviews/approves it before submission to report preparer (e.g., Institutional Research) for reporting, etc.*

(2) Carry-out and implement unit-level controls.

(3) Maintain evidence of unit-level controls.

(2) Carry-out validation, verification, and security controls:

	Data Milestone	Examples of controls – validation, verification, and security
(1)	Before, during and after data input/capture	(a) Control access to spreadsheets and data systems (b) Enable system/spreadsheet audit logs to track changes (c) System/spreadsheet edit checks (prevent wrong data element type, range, etc.) (d) Designate independent reviewers to confirm correct capture (e) For automated feeds, establish a transfer log/record count check
(2)	After extraction from storage	(f) Perform reconciliation – manual or automated - between data extracted versus data source (entire population or sampling)
(3)	After compilation for reporting purposes	(g) Agree on context and usage of data (Institutional Research and data owner/supplier) (h) Retain supporting data sets/results that substantiate reported values (i) Establish documented unit-level (data owner/supplier) *data verification and approval process,* which is relied upon by the report preparer (e.g., Institutional Research, who performs a direct query of your data OR receives a spreadsheet/data file from your unit).
(4)	Published data	(j) Perform a post-publication review that data published was the data agreed upon.

(3) Maintain evidence of unit-level data validation, verification, and approval process

Examples of evidence include:

- Unit-level approval support (e.g., emails string from designated department level reviewers/approvers)
- Hardcopy sign-off approvals on forms maintained, etc.
- Approvals should be supported by retention of query and results supporting the submitted data.

Questions?